

Convergence Rates for the Variable,  
the Multiplier, and the Pair  
in SQP Methods

Jershan Chiang and Richard Tapia

March, 1989  
(revised June 1989)

TR89-2

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE <b>MAR 1989</b>		2. REPORT TYPE		3. DATES COVERED <b>00-00-1989 to 00-00-1989</b>	
4. TITLE AND SUBTITLE <b>Convergence Rates for the Variable, the Multiplier, and the Pair in SQP Methods</b>				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892</b>				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES <b>34</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

# Convergence Rates for the Variable, the Multiplier, and the Pair in SQP Methods<sup>1,2</sup>

Jershan Chiang<sup>3</sup>

Richard Tapia<sup>4</sup>

March 1989

## Abstract

This work investigates relationships among the convergence rates for the variable  $x$ , for the multiplier  $\lambda$  and for the pair  $(x, \lambda)$  in SQP methods for equality constrained optimization. Key contributions are: if the convergence in  $(x, \lambda)$  and also in  $x$  is  $q$ -superlinear, then the convergence in  $\lambda$  is either  $q$ -superlinear or  $q$ -sublinear with unbounded  $q_1$  factor, and if the convergence in  $(x, \lambda)$  is  $q$ -superlinear, then the convergence in  $x$  is at least two-step  $q$ -superlinear. It is noted that a theorem of Fontecilla, Steihaug and Tapia leads to a characterization result which is potentially more useful than the Boggs-Tolle-Wang characterization. Finally, two different conditions that guarantee  $q$ -superlinear convergence in  $x$ ,  $\lambda$  and  $(x, \lambda)$  for an SQP method are derived.

**Key words:** secant, quasi-Newton, SQP method, superlinear convergence  
1980 Mathematics Subject Classification: Primary 49D15, 65K05

---

<sup>1</sup>Research sponsored by SDIO/IST/ARO, AFOSR 85-0243, and DOE DEFG05-86ER25017.

<sup>2</sup>A portion of this work is contained in the first author's doctoral thesis under the supervision of the second author in the Department of Mathematical Sciences, Rice University.

<sup>3</sup>Department of Applied Mathematics, Chinese Culture University, Taipei, Taiwan.

<sup>4</sup>Department of Mathematical Sciences, Rice University, Houston, Texas 77251-1892.



# 1 Introduction

In this work we will be concerned with the equality constrained optimization problem

$$\begin{aligned} & \text{minimize} && f(x) \\ & \text{subject to} && h_i(x) = 0, \quad i = 1, 2, \dots, m \end{aligned} \quad (1.1)$$

where  $f, h_i$  are nonlinear functions defined from  $\mathbb{R}^n$  into  $\mathbb{R}$ .

We denote by  $h(x)$  the vector whose components are  $h_i(x)$ ,  $i = 1, \dots, m$ . The *Lagrangian function* associated with problem (1.1) is the function

$$\ell(x, \lambda) = f(x) + \lambda^T h(x) \quad (1.2)$$

where  $\lambda = (\lambda_1, \dots, \lambda_m)^T$  is called the vector of *Lagrange multipliers* or simply the *Lagrange multiplier*. The *augmented Lagrangian function* associated with problem (1.1) is the function

$$L(x, \lambda; \rho) = f(x) + \lambda^T h(x) + \frac{1}{2} \rho h(x)^T h(x) \quad (\rho \geq 0). \quad (1.3)$$

The algorithm we are interested in is the successive quadratic programming (SQP) Lagrangian quasi-Newton method:

ALGORITHM (SQP Method):

For  $k = 0, 1, \dots$ , until convergence do

$$x_{k+1} = x_k + s_k \quad (1.4)$$

$$\lambda_{k+1} = \lambda_k + \Delta \lambda_k \quad (1.5)$$

$$B_{k+1} = \mathbb{B}(x_k, s_k, \lambda_{k+1}, B_k) \quad (1.6)$$

where  $s_k$  and  $\Delta \lambda_k$  are the solution and the multiplier associated with the solution of the quadratic program

$$\begin{aligned} & \text{minimize} && \nabla_x \ell(x_k, \lambda_k)^T s + \frac{1}{2} s^T B_k s \\ & \text{subject to} && \nabla h(x_k)^T s + h(x_k) = 0. \end{aligned} \quad (1.7)$$

The matrix  $B_{k+1}$  is interpreted as an approximation to  $\nabla_x^2 \ell(x_{k+1}, \lambda_{k+1})$ . When the augmented Lagrangian is substituted for the Lagrangian in the SQP method we call the resulting algorithm the SQP augmented Lagrangian

quasi-Newton method. For further details on these various SQP formulations see Appendix A of Tapia (1988).

We begin our study with a review of terminology concerning convergence rates for iterative methods. For the most part we follow Chapter 9 of Ortega and Rheinboldt (1970). However, our definition of  $r$ -convergence is essentially that of Dennis and Schnabel (1983), which is known to be equivalent to the notion considered by Ortega and Rheinboldt (1970).

Let  $\{x_k\} \subset \mathbb{R}^n$  be a convergent sequence with limit  $x_*$  and assume that  $x_k \neq x_*$  for all  $k$ . Consider a vector norm  $\|\cdot\|$  on  $\mathbb{R}^n$ . for  $p \in [1, \infty)$  the quantities

$$q_p = \overline{\lim}_k \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|^p}$$

are called the  $q_p$  factors of the sequence  $\{x_k\}$  with respect to the norm  $\|\cdot\|$ .

We define the  $q$ -order of convergence of  $\{x_k\}$  to be  $\inf\{p : q_p = \infty\}$ . The  $q_1$  factor will be of particular interest to us. If  $q_1 < 1$ , then the convergence is said to be  $q$ -linear; while if  $q_1 \geq 1$ , then the convergence is said to be  $q$ -sublinear. Clearly the ideal situation is when  $q_1 = 0$  and in this case the convergence is said to be  $q$ -superlinear. The least ideal case is when  $q_1 = +\infty$ . We will refer to this convergence as  $q$ -sublinear with unbounded  $q_1$  factor.

Suppose that we have  $\{b_k\}$  converging to zero and such that  $\|x_k - x_*\| \leq b_k$  for all  $k$ . If the sequence  $\{b_k\}$  possesses a particular  $q$ -convergence property, then the sequence  $\{x_k\}$  is said to possess the corresponding  $r$ -convergence property.

If for each  $k$  the subsequence  $x_k, x_{k+j}, x_{k+2j}, \dots$  displays a particular convergence behavior, then we say that the original sequence  $x_1, x_2, \dots$  has this  $j$ -step convergence behavior.

It is of interest to observe that  $r$ -convergence properties are norm independent; so are the notions of  $q$ -order,  $q$ -superlinear and  $q$ -sublinear with unbounded  $q_1$  factor. However, the notions of  $q$ -linear and  $q$ -sublinear are norm dependent.

Convergence of order 2 is said to be *quadratic* and that of order 3 is said to be *cubic*. Unfortunately this standard terminology is such that a  $q$ -order of 1 does not imply  $q$ -linear convergence.

The SQP Lagrangian method for equality constrained optimization is a part of the optimization theory folklore. It was certainly known to researchers in the calculus of variations in the early 1900's. The fact that the

SQP Lagrangian Newton method, i.e.  $B_k = \nabla_x^2 \ell(x_k, \lambda_k)$ , is  $q$ -quadratically convergent in the pair  $(x, \lambda)$  is also a part of the optimization folklore. For a proof see Tapia (1977). The SQP Lagrangian Newton method for problems with inequality constraints is usually credited to Wilson (1963).

The convergence of special SQP secant methods has been investigated by many authors. The first was probably Garcia-Palomares and Mangasarian (1976). They posed an SQP secant method and proved various  $r$ -convergence results in  $(x, \lambda)$  under certain assumptions. Han (1976), (1977), Tapia (1977) and Glad (1979) independently established local and  $q$ -superlinear convergence in  $(x, \lambda)$  for several SQP secant methods. Boggs, Tolle and Wang (1982) showed that the convergence in  $x$  for the SQP, DFP and BFGS secant method was  $q$ -superlinear assuming that  $\{x_k\}$  converged  $q$ -linearly. They obtained a characterization of  $q$ -superlinear convergence in  $x$  also assuming that  $\{x_k\}$  converged  $q$ -linearly. Fontecilla, Steihaug and Tapia (1987) established  $q$ -superlinear convergence in  $x$  for the SQP PSB, DFP and BFGS secant methods and also derived the Boggs-Tolle-Wang characterization of  $q$ -superlinear convergence in  $x$  as a special case of a more general characterization. These results did not require the assumption that  $\{x_k\}$  converged  $q$ -linearly.

When the secant update in question was the DFP or the BFGS update, all of the above authors were forced to either make the assumption that the Hessian with respect to  $x$  of the Lagrangian at the solution was positive definite or work with the SQP augmented Lagrangian method with the penalty constant chosen sufficiently large. Each one of these two alternatives is somewhat undesirable. The first because the Hessian with respect to  $x$  of the Lagrangian is not in general positive definite at the solution, and the second because the SQP augmented Lagrangian method is known to be sensitive to the choice of the penalty constant  $\rho$ , and adequate guidelines for this choice seem to be impossible to develop.

This unfortunate state of affairs motivated Powell (1978) to propose an *ad hoc* modification to the SQP Lagrangian BFGS secant method which compensates for the lack of positive definiteness in the Hessian at the solution. Assuming convergence, Powell was able to show that his modified SQP BFGS secant method gave  $r$ -superlinear convergence in the variable  $x$ .

The dilemma described above, i.e. the lack of positive definiteness of the Hessian with respect to  $x$  of the Lagrangian at the solution, motivated considerable research activity in formulating a BFGS secant method for problem

(1.1) in the framework of the so-called reduced Hessian methods. In contrast to full Hessian methods, the reduced Hessian methods approximate the Hessian restricted to a proper subspace where it is expected to be positive definite. For more on reduced Hessian secant methods see Murray and Wright (1978), Gabay (1982), Coleman and Conn (1984), Nocedal and Overton (1985), Gurwitz (1986), Gilbert (1987), and Byrd and Nocedal (1988). The theoretical convergence rate that has been obtained for the various reduced Hessian BFGS secant methods is two-step  $q$ -superlinear convergence in the variable  $x$ .

Fenyés (1987) and Fontecilla (1988) propose full Hessian methods which have some of the flavor of the reduced Hessian methods.

Recently Tapia (1988) proposed two new classes of SQP secant methods for problem (1.1). One class consists of SQP Lagrangian secant methods with a modification in the scale associated with the particular secant update in question to compensate for the lack of positive definiteness of the Hessian with respect to  $x$  of the Lagrangian. The other class consists of SQP structured augmented Lagrangian secant methods. From an algorithmic point of view, these methods possess the flavor of the Powell modified SQP BFGS secant method. However, Tapia was able to prove that for both methods the DFP and the BFGS versions of the algorithms are locally convergent and give  $q$ -superlinear convergence in the pair  $(x, \lambda)$  and also in the variable  $x$  without a positive definite assumption on the Hessian with respect to  $x$  of the Lagrangian. Current research activity is attempting to demonstrate that satisfactory rules exist for choosing the parameter in these methods that corresponds to the penalty parameter in the Hessian with respect to  $x$  of the augmented Lagrangian.

To our knowledge there are no results in the literature that concern the  $q$ -convergence rate of the multiplier in an SQP method. However, several recent articles in the literature made various assumptions concerning this rate. For example, Boggs and Tolle (1985) considered the SQP Lagrangian BFGS and DFP secant methods. They showed that the convergence in  $x$  was  $q$ -superlinear assuming that the convergence in  $x$  and  $\lambda$  was  $q$ -linear and  $\{x_k\}$  satisfied a condition called tangential convergence. This result did not require the Hessian of the Lagrangian with respect to  $x$  to be positive definite at the solution. Gill, Murray, Saunders and Wright (1986) proposed an SQP Lagrangian secant method for generating a search direction and determined the step length from a line-search strategy with an augmented Lagrangian



as the merit function. They established, under the assumptions that the convergence in  $x$  and in  $\lambda$  were  $q$ -superlinear and  $\frac{\|\Delta x_k\|}{\|\Delta \lambda_k\|} > M > 0$  for  $k$  sufficiently large, that eventually their steplength choice would be 1.

Recently Tapia and Whitley (1988) investigated the convergence rate of the projected Newton method applied to the symmetric eigenvalue problem. This algorithm can be viewed as the SQP Newton method followed by a two-norm normalization. They established a  $q$ -rate of convergence of  $1 + \sqrt{2}$  for both the variable  $x$  and the multiplier  $\lambda$ .

The result of Tapia and Whitley (1988) further motivated and strengthened our already strong desire to determine relationships among the convergence rates for  $(x, \lambda)$ , for  $x$  and for  $\lambda$  in SQP methods.

Section 2 deals with preliminaries including our notation and assumptions. In Section 3 we collect three useful characterization theorems. The first result is a straightforward application of the well-known Dennis-Moré characterization. We derive it in a very convenient form which readily lends itself to applications. The second is the Boggs-Tolle-Wang characterization and the third is a useful characterization theory which follows directly by restricting a theorem of Fontecilla, Steihaug and Tapia to the case of SQP. We maintain that collectively these three theorems offer a powerful tool and demonstrate this fact by using them to derive some interesting consequences. Two conditions which guarantee the  $q$ -superlinear convergence of the multiplier sequence  $\{\lambda_k\}$  are included. However, we wish to emphasize that the main theme of this work is that while the  $q$ -superlinear convergence of  $\{\lambda_k\}$  is likely in any given computation, mathematical conditions which assume it or guarantee it are necessarily restrictive.

In Section 4 we show that the convergence in  $x$  is at least two-step  $q$ -superlinear whenever the convergence in  $(x, \lambda)$  is  $q$ -superlinear. In Section 5 we show that if the convergence is  $(x, \lambda)$  and  $x$  is  $q$ -superlinear, then the convergence in  $\lambda$  is either  $q$ -superlinear or  $q$ -sublinear with unbounded  $q_1$  factor. We consider these two theorems to be key contributions of the paper. In Section 6 we summarize and present some concluding remarks.

## 2 Preliminaries

In an effort to simplify our notation and give a cleaner presentation we will work exclusively with the SQP Lagrangian formulation in the remainder of this work, i.e., we will effectively choose  $\rho = 0$  in the augmented Lagrangian. No loss of generality will result from this simplification as long as we remember that the requirement that the Hessian of the Lagrangian with respect to  $x$  be positive definite at the solution, can be dealt with by working with the augmented Lagrangian and choosing  $\rho$  sufficiently large.

Let  $x_*$  be a local solution of Problem (1.1) with associated multiplier  $\lambda_*$ . We will use the notation  $\nabla h_k = \nabla h(x_k)$ ,  $\nabla f_k = \nabla f(x_k)$ ,  $\nabla h_* = \nabla h(x_*)$  and  $A_* = \nabla_x^2 \ell(x_*, \lambda_*)$ . Both the  $\ell_2$  vector-norm and the corresponding induced matrix norm will be denoted by  $|\cdot|$ . We will use  $\|\cdot\|$  to denote an arbitrary but fixed matrix norm.

Throughout this work we make the following assumptions:

- A1. The functions  $f$  and  $h$  have continuous second derivatives in an open neighborhood  $D$  of a local solution  $x_*$  of problem (1.1) and these second derivatives are Lipschitz continuous at  $x_*$ .
- A2.  $\nabla h_*$  has full rank.
- A3.  $z^T A_* z > 0$  for all  $z \neq 0$  satisfying  $\nabla h_*^T z = 0$ .
- A4. For large  $k$  the sequence  $\{(x_k, \lambda_k)\}$  has been generated by a particular SQP quasi-Newton method with invertible  $B_k$ . Also  $\{(x_k, \lambda_k)\}$  converges to  $(x_*, \lambda_*)$ .

Assumptions A1, A2, and A3 are standard assumptions in the study of quasi-Newton methods for constrained optimization. Assumption A3 is the well-known second-order sufficiency condition.

Since we are only concerned with convergence rates, no generality will be lost and considerable simplicity will be gained by assuming that assumption A4 holds for all  $k$  and from assumptions A1, A2, and A4 that  $\nabla h_k$  has full rank for all  $k$ .

The requirement that  $B_k$  be invertible is mild and is effectively implied by second-order sufficiency for the subproblem (1.7). To see this observe that if  $B_k$  is positive definite on  $\{\eta : \nabla h_k^T \eta = 0\}$ , then for sufficiently large  $\rho$

the matrix  $\hat{B}_k = B_k + \rho \nabla h_k \nabla h_k^T$  is positive definite and therefore invertible. Moreover, the subproblem (1.7) has the same solution using  $B_k$  or  $\hat{B}_k$ . For more detail see Tapia (1977).

Some of our results will also require that  $\{B_k\}$  and  $\{B_k^{-1}\}$  be bounded. The requirement that  $\{B_k\}$  and  $\{B_k^{-1}\}$  be bounded is quite mild. In general if the quasi-Newton update satisfies bounded deterioration, then we have that  $\{B_k\}$  and  $\{B_k^{-1}\}$  are bounded if the initial  $(x_0, B_0)$  is close to  $(x_*, A_*)$ . Moreover, the well-known secant updates Broyden, PSB, DFP and BFGS all satisfy bounded deterioration. For more details see the argument used in Broyden, Dennis and Moré (1973) and Theorem 3.1 and 3.2 in Fontecilla, Steihaug and Tapia (1987).

In our study we will have need to refer to the SQP quasi-Newton method in its equivalent “diagonalized multiplier method” form

$$\lambda_{k+1} = (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} (h_k - \nabla h_k^T B_k^{-1} \nabla f_k) \quad (2.1)$$

and

$$x_{k+1} = x_k - B_k^{-1} \nabla_x \ell(x_k, \lambda_{k+1}) . \quad (2.2)$$

For details on this equivalence see Tapia (1977), (1978) or Fontecilla, Steihaug and Tapia (1987).

In many of our results we will need to relate the quantity  $\lambda_{k+1} - \lambda_*$  to the quantities  $x_{k+1} - x_*$ , and  $x_k - x_*$ . The following lemmas are technical results which accomplish this objective and will be useful tools in the proof of several of our results.

**Lemma 2.1** *There exists a sequence of matrices  $\{\Gamma_k\}$  such that  $\{\Gamma_k\}$  converges to  $A_*$  and*

$$x_{k+1} - x_* + B_k^{-1} \nabla h_k (\lambda_{k+1} - \lambda_*) = (I - B_k^{-1} \Gamma_k) (x_k - x_*) .$$

*Proof.* From (2.2) we can write

$$\begin{aligned} x_{k+1} - x_* &= x_k - x_* - B_k^{-1} [\nabla_x \ell(x_k, \lambda_{k+1}) - \nabla_x \ell(x_*, \lambda_*)] \\ &= x_k - x_* - B_k^{-1} [\nabla f_k + \nabla h_k \lambda_{k+1} - \nabla f_* - \nabla h_* \lambda_*] \\ &= B_k^{-1} [B_k(x_k - x_*) - (\nabla f_k - \nabla f_*) - \nabla h_k (\lambda_{k+1} - \lambda_*) \\ &\quad - (\nabla h_k - \nabla h_*) \lambda_*] \end{aligned}$$

$$\begin{aligned}
&= B_k^{-1} \left\{ B_k(x_k - x_*) - \left[ \int_0^1 \nabla^2 f(x_* + t(x_k - x_*)) dt \right] (x_k - x_*) \right. \\
&\quad \left. - \left[ \int_0^1 \nabla^2 h(x_* + t(x_k - x_*)) \lambda_* dt \right] (x_k - x_*) \right\} \\
&\quad - B_k^{-1} \nabla h_k(\lambda_{k+1} - \lambda_*) ,
\end{aligned}$$

where the integral of the matrix-valued function is interpreted component-wise. For more details see Chapter 4 of Dennis and Schnabel (1983). Let  $\Gamma_k = \int_0^1 \nabla^2 f(x_* + t(x_k - x_*)) dt + \int_0^1 \nabla^2 h(x_* + t(x_k - x_*)) \lambda_* dt$ . Then we have

$$x_{k+1} - x_* = B_k^{-1} [B_k(x_k - x_*) - \Gamma_k(x_k - x_*)] - B_k^{-1} \nabla h_k(\lambda_{k+1} - \lambda_*) ,$$

or

$$x_{k+1} - x_* + B_k^{-1} \nabla h_k(\lambda_{k+1} - \lambda_*) = (I - B_k^{-1} \Gamma_k)(x_k - x_*) .$$

By the definition of  $\Gamma_k$  and the fact that  $\{(x_k, \lambda_k)\}$  converges to  $(x_*, \lambda_*)$  we have  $\{\Gamma_k\}$  converges to  $A_*$ .  $\square$

**Lemma 2.2** *We have*

$$\lambda_{k+1} - \lambda_* = (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} [\nabla h_k^T (I - B_k^{-1} \Gamma_k)(x_k - x_*) + O(|x_k - x_*|^2)] ,$$

where  $\{\Gamma_k\}$  is as in Lemma 2.1.

*Proof.* From (2.1) we can write

$$\begin{aligned}
\lambda_{k+1} &= \lambda_k + (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} [h_k - \nabla h_k^T B_k^{-1} \nabla_x \ell(x_k, \lambda_k)] \\
&= \lambda_k + (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} \{h_k - h_* - \nabla h_k^T B_k^{-1} [\nabla_x \ell(x_k, \lambda_k) \\
&\quad - \nabla_x \ell(x_*, \lambda_*)]\} .
\end{aligned}$$

Define  $\Gamma_k$  by the same formula used in Lemma 2.1 and perform the same algebra to obtain

$$\begin{aligned}
\lambda_{k+1} &= \lambda_k + (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} \left\{ \left[ \int_0^1 \nabla h(x_* + t(x_k - x_*))^T dt \right] (x_k - x_*) \right. \\
&\quad \left. - \nabla h_k^T B_k^{-1} \Gamma_k(x_k - x_*) \right\} - (\lambda_k - \lambda_*) .
\end{aligned}$$

Hence

$$\lambda_{k+1} - \lambda_* = (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} [\nabla h_k^T (I - B_k^{-1} \Gamma_k)(x_k - x_*) + O(|x_k - x_*|^2)] .$$

□

Consider  $a_k \geq 0$  and  $b_k > 0$ . We write  $a_k = O(b_k)$  if there exists  $m$  such that  $\frac{a_k}{b_k} \leq m$  for all  $k$ . We now use Lemma 2.2 to establish that  $|\lambda_{k+1} - \lambda_*| = O(|x_k - x_*|)$  whenever  $\{B_k\}$  and  $\{B_k^{-1}\}$  are bounded.

**Theorem 2.1** *If  $\{B_k\}$  and  $\{B_k^{-1}\}$  are bounded then  $|\lambda_{k+1} - \lambda_*| = O(|x_k - x_*|)$ .*

*Proof.* From Lemma 2.2 we have

$$\lambda_{k+1} - \lambda_* = (\nabla h_k^T B_k^{-1} \nabla h_k)^{-1} [\nabla h_k^T (I - B_k^{-1} \Gamma_k)(x_k - x_*) + O(|x_k - x_*|^2)] . \quad (2.3)$$

From Lemma 3.7 in Fontecilla (1988) we have

$$|(\nabla h_k^T B_k^{-1} \nabla h_k)^{-1}| \leq |(\nabla h_k^T \nabla h_k)^{-1} \nabla h_k^T|^2 |B_k| .$$

Since  $\{B_k\}$  is bounded and  $\nabla h_*$  has full rank, it follows that there exists  $C_1$  such that

$$|(\nabla h_k^T B_k^{-1} \nabla h_k)^{-1}| \leq C_1 . \quad (2.4)$$

Furthermore

$$\begin{aligned} |(\nabla h_k^T (I - B_k^{-1} \Gamma_k))| &= |\nabla h_k^T B_k^{-1} (B_k - \Gamma_k)| \\ &\leq |\nabla h_k^T| |B_k^{-1}| |B_k - \Gamma_k| . \end{aligned}$$

The facts that  $\{B_k^{-1}\}$  is bounded and  $\{\Gamma_k\}$  converges to  $A_*$  lead to

$$|\nabla h_k^T (I - B_k^{-1} \Gamma_k)| \leq C_2 \quad (2.5)$$

for some constant  $C_2$ .

Combining (2.3), (2.4) and (2.5) we have  $\frac{|\lambda_{k+1} - \lambda_*|}{|x_k - x_*|} \leq C_3$  for some constant  $C_3$ . Therefore  $|\lambda_{k+1} - \lambda_*| = O(|x_k - x_*|)$ . □

Using Theorem 2.1 we can establish the finiteness of the  $q_1$  factor of the sequence  $\{x_k\}$ .

**Corollary 2.1** *If  $\{B_k\}$  and  $\{B_k^{-1}\}$  are bounded, then*

$$|x_{k+1} - x_*| = O(|x_k - x_*|) .$$

*Proof.* By Lemma 2.1 we have

$$x_{k+1} - x_* - (I - B_k^{-1}\Gamma_k)(x_k - x_*) = -B_k^{-1}\nabla h_k(\lambda_{k+1} - \lambda_*) . \quad (2.6)$$

From Theorem 2.1 and the fact that  $\{B_k^{-1}\}$  is bounded we have

$$\begin{aligned} \frac{|B_k^{-1}\nabla h_k(\lambda_{k+1} - \lambda_*)|}{|x_k - x_*|} &\leq |B_k^{-1}| |\nabla h_k| \frac{|\lambda_{k+1} - \lambda_*|}{|x_k - x_*|} \\ &\leq C_1 \end{aligned} \quad (2.7)$$

for some constant  $C_1$ . Furthermore

$$\begin{aligned} \frac{|(I - B_k^{-1}\Gamma_k)(x_k - x_*)|}{|x_k - x_*|} &= \frac{|B_k^{-1}(B_k - \Gamma_k)(x_k - x_*)|}{|x_k - x_*|} \\ &\leq |B_k^{-1}| |B_k - \Gamma_k| . \end{aligned}$$

By assumption  $\{B_k^{-1}\}$  is bounded, hence

$$\frac{|(I - B_k^{-1}\Gamma_k)(x_k - x_*)|}{|x_k - x_*|} \leq C_2 \quad (2.8)$$

for some constant  $C_2$ . Combining (2.6), (2.7), and (2.8) we have

$$|x_{k+1} - x_*| = O(|x_k - x_*|) . \quad \square$$

The following result holds for any sequences  $\{x_k\}$  and  $\{\lambda_k\}$  no matter whether they were generated by an SQP quasi-Newton method or not.

**Proposition 2.1** *If  $\{x_k\}$  converges to  $x_*$   $q$ -superlinearly and  $\{\lambda_k\}$  converges to  $\lambda_*$   $q$ -superlinearly, then  $\{(x_k, \lambda_k)\}$  converges to  $(x_*, \lambda_*)$   $q$ -superlinearly.*

*Proof.* Since  $q$ -superlinear convergence is independent of norm, we can work with the max norm.

Let  $\|\cdot\|$  denote the max norm. By assumption there exist  $\{c_k\}$  and  $\{\tilde{c}_k\}$  such that

$$\begin{aligned} \text{and} \quad \|x_{k+1} - x_*\| &\leq c_k \|x_k - x_*\| \\ \|\lambda_{k+1} - \lambda_*\| &\leq \tilde{c}_k \|\lambda_k - \lambda_*\| \end{aligned}$$

where  $\{c_k\}$  and  $\{\tilde{c}_k\}$  converge to 0. Then

$$\begin{aligned}
\|(x_{k+1}, \lambda_{k+1}) - (x_*, \lambda_*)\| &= \|(x_{k+1} - x_*, \lambda_{k+1} - \lambda_*)\| \\
&= \max\{\|x_{k+1} - x_*\|, \|\lambda_{k+1} - \lambda_*\|\} \\
&\leq \max\{c_k\|x_k - x_*\|, \tilde{c}_k\|\lambda_k - \lambda_*\|\} \\
&\leq (\max\{c_k, \tilde{c}_k\}) \|(x_k - x_*, \lambda_k - \lambda_*)\|.
\end{aligned}$$

It follows that  $\{(x_k, \lambda_k)\}$  converges to  $(x_*, \lambda_*)$   $q$ -superlinearly.  $\square$

### 3 Fundamental Characterization Theorems and Consequences

In this section we present three important characterization theorems and then derive several consequences of these theorems. One of these theorems, the Boggs-Tolle-Wang characterization, is somewhat well-known. Another one of these theorems results from a straightforward application of the well-known Dennis-Moré characterization. We state it in a form which is particularly convenient. The remainder of the results presented seem to be unknown.

It is straightforward that the SQP method can be viewed as a quasi-Newton method applied to the nonlinear equations which represent the first-order necessary conditions for problem (1.1); see Tapia (1977), (1978) for details. Applications of the Dennis-Moré characterization of  $q$ -superlinear convergence for quasi-Newton methods have been used in this context by numerous authors. We now use the Dennis-Moré characterization to derive the following characterization.

We write  $\Delta x_k$  for  $x_{k+1} - x_k$  and similarly for  $\Delta \lambda_k$ .

**Theorem 3.1** *For an SQP method the following two statements are equivalent:*

- (i)  $\{(x_k, \lambda_k)\}$  converges  $q$ -superlinearly to  $(x_*, \lambda_*)$ .
- (ii)  $\lim_{k \rightarrow \infty} \left\{ \frac{|(B_k - A_*)\Delta x_k|}{|\Delta x_k|} \min \left( 1, \frac{|\Delta x_k|}{|\Delta \lambda_k|} \right) \right\} = 0$

*Proof.* To start with observe that both properties (i) and (ii) are norm independent. Hence we may conveniently work with the max norm, which we denote by  $\|\cdot\|$ .

According to the Dennis-Moré superlinear convergence criterion (see Dennis-Moré (1974)) we have  $q$ -superlinear convergence of the sequence  $\{(x_k, \lambda_k)\}$  if and only if

$$\lim_{k \rightarrow \infty} \frac{\|(B_k - A_*)\Delta x_k + (\nabla h_k - \nabla h_*)\Delta \lambda_k\|}{\|(\Delta x_k, \Delta \lambda_k)\|} = 0 \quad (3.1)$$

and

$$\lim_{k \rightarrow \infty} \frac{\|(\nabla h_k - \nabla h_*)^T \Delta x_k\|}{\|(\Delta x_k, \Delta \lambda_k)\|} = 0. \quad (3.2)$$

See equations (129)–(131) of Tapia (1977) for further details. Clearly (3.2) holds and (3.1) is equivalent to

$$\lim_{k \rightarrow \infty} \frac{\|(B_k - A_*)\Delta x_k\|}{\|(\Delta x_k, \Delta \lambda_k)\|} = 0. \quad (3.3)$$

Dividing (3.3) by  $\|\Delta x_k\|$  we see that (3.3) is equivalent to condition (ii). The Dennis-Moré characterization requires  $\nabla^2 \ell(x_*, \lambda_*)$  to be nonsingular. This follows from A2 and A3.  $\square$

The following is the well-known Boggs-Tolle-Wang characterization. For a self-contained and short proof see Stoer and Tapia (1987).

Let

$$P_k = I - \nabla h_k (\nabla h_k^T \nabla h_k)^{-1} \nabla h_k^T. \quad (3.4)$$

**Theorem 3.2** (*Boggs-Tolle-Wang*). *For an SQP method the following two statements are equivalent:*

(i)  $\{x_k\}$  converges  $q$ -superlinearly to  $x_*$ .

(ii)  $\lim_{k \rightarrow \infty} \frac{|P_k(B_k - A_*)\Delta x_k|}{|\Delta x_k|} = 0$ .

The SQP Broyden, PSB, DFP and BFGS secant methods are known to satisfy the condition

$$\lim_{k \rightarrow \infty} \frac{|(B_k - A_*)\Delta x_k|}{|\Delta x_k|} = 0, \quad (3.5)$$



hence by Theorem 3.1 and Theorem 3.2 they give  $q$ -superlinear convergence in  $x$  and in  $(x, \lambda)$ . For details see Corollary 5.5 in Fontecilla, Steihaug and Tapia (1987).

These comments bring to the foreground the following common concern. Many researchers in the area find it unsettling that in every known situation where the  $q$ -superlinear convergence of  $x$  has been established for an SQP method, it was established by first demonstrating that (3.5) holds. Clearly (3.5) implies the Boggs-Tolle-Wang condition ((ii) of Theorem 3.2). This means that we are sacrificing information by using the Boggs-Tolle-Wang characterization, since we must have more than just  $q$ -superlinear convergence in  $x$ . Of course, the pertinent issue here is a characterization of condition (3.5) in terms of the convergence aspects of the SQP method in question. We find it interesting that such a characterization has not been identified and yet it can be obtained as a straightforward consequence of Theorem 5.1 of Fontecilla, Steihaug and Tapia (1987). Specifically, by restricting their theorem to the case of SQP and making some rather obvious observations we obtain the following result.

**Theorem 3.3** (*Fontecilla-Steihaug-Tapia*). *For an SQP method the following two statements are equivalent:*

- (i)  $\lim_{k \rightarrow \infty} \frac{|(B_k - A_*)\Delta x_k|}{|\Delta x_k|} = 0.$
- (ii) (a)  $\{x_k\}$  converges  $q$ -superlinearly to  $x_*$  and  
(b)  $\lim_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|x_k - x_*|} = 0.$

We now present several consequences of Theorem 3.1 and Theorem 3.3. Before we isolated these two theorems we had derived these results using only the Dennis-Moré characterization. This was a lengthy task. Hence these consequences also can be viewed as a demonstration that collectively these theorems offer a powerful analytical tool.

**Proposition 3.1** *For an SQP method the statement*

$$(i) \lim_{k \rightarrow \infty} \frac{|(B_k - A_*)\Delta x_k|}{|\Delta x_k|} = 0$$

or the statement

- (ii) (a)  $\{B_k\}$  is bounded and
- (b)  $\lim_{k \rightarrow \infty} \frac{|\Delta x_k|}{|\Delta \lambda_k|} = 0$

implies the statement

- (iii)  $\{(x_k, \lambda_k)\}$  converges  $q$ -superlinearly to  $(x_*, \lambda_*)$ .

*Proof.* The proof follows directly from Theorem 3.1.  $\square$

We now extend the Boggs-Tolle-Wang characterization of  $q$ -superlinear convergence of  $\{x_k\}$  to the common case where it is known that the pair  $(x, \lambda)$  converges  $q$ -superlinearly.

**Proposition 3.2** *For an SQP method assume that  $\{(x_k, \lambda_k)\}$  converges  $q$ -superlinearly to  $(x_*, \lambda_*)$ . Then the following two statements are equivalent:*

- (i)  $\{x_k\}$  converges  $q$ -superlinearly to  $x_*$ .

- (ii)

$$\lim_{k \rightarrow \infty} \frac{|P_{n_k}(B_{n_k} - A_*)\Delta x_{n_k}|}{|\Delta x_{n_k}|} = 0 \quad (3.6)$$

whenever  $\{n_k\}$  is a sequence of positive integers such that

$$\lim_{k \rightarrow \infty} \frac{|\Delta x_{n_k}|}{|\Delta \lambda_{n_k}|} = 0. \quad (3.7)$$

*Proof.* By Theorem 3.2 (i) implies (ii). Now suppose that

$$\overline{\lim}_{k \rightarrow \infty} \frac{|P_k(B_k - A_*)\Delta x_k|}{|\Delta x_k|} = \delta > 0. \quad (3.8)$$

Then for some sequence of positive integer  $\{n_k\}$

$$\lim_{k \rightarrow \infty} \frac{|P_{n_k}(B_{n_k} - A_*)\Delta x_{n_k}|}{|\Delta x_{n_k}|} = \delta > 0.$$

It follows that

$$\overline{\lim}_{k \rightarrow \infty} \frac{|(B_{n_k} - A_*)\Delta x_{n_k}|}{|\Delta x_{n_k}|} = \delta' > 0.$$

Choosing a subsequence of  $\{n_k\}$ , and again calling it  $\{n_k\}$ , we have

$$\lim_{k \rightarrow \infty} \frac{|(B_{n_k} - A_*)\Delta x_{n_k}|}{|\Delta x_{n_k}|} = \delta' > 0. \quad (3.9)$$

From Theorem 3.1 we see that (3.9) implies (3.7); hence (3.6). However (3.6) and (3.9) are incompatible. It follows that our supposition (3.8) cannot hold and (ii) of Theorem 3.2 must hold.  $\square$

Observe that (ii) vacuously holds if

$$\frac{|\Delta x_k|}{|\Delta \lambda_k|} \geq m > 0 \quad \text{for all } k. \quad (3.10)$$

Hence (3.10) and the  $q$ -superlinear convergence of  $\{(x_k, \lambda_k)\}$  imply the  $q$ -superlinear convergence of  $\{x_k\}$ .

**Proposition 3.3** *For an SQP method assume the condition*

$$\frac{|x_k - x_*|}{|\lambda_k - \lambda_*|} \leq M < +\infty \quad \text{for all } k. \quad (3.11)$$

*Then the statement*

- (i)  $\lim_{k \rightarrow \infty} \frac{|(B_k - A_*)\Delta x_k|}{|\Delta x_k|} = 0$  *implies the statement*
- (ii) (a)  $\{x_k\}$  *converges*  $q$ -*superlinearly to*  $x_*$  *and*  
 (b)  $\{\lambda_k\}$  *converges*  $q$ -*superlinearly to*  $\lambda_*$ .

*Conversely, if instead of (3.11) we assume*

$$0 < m \leq \frac{|x_k - x_*|}{|\lambda_k - \lambda_*|} \quad \text{for all } k, \quad (3.12)$$

*then the statement (ii) implies the statement (i). Consequently assuming (3.11) and (3.12) we have that (i) is equivalent to (ii).*

*Proof.* Consider the expression

$$\frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} = \frac{|\lambda_{k+1} - \lambda_*|}{|x_k - x_*|} \frac{|x_k - x_*|}{|\lambda_k - \lambda_*|}. \quad (3.13)$$

If (3.11) holds, then Theorem 3.3, statement (i) and (3.13) imply (ii). Now, if (3.12) holds, then Theorem 3.3, statement (ii) and (3.13) imply (i) holds.  $\square$

We believe that in most cases (i) and (ii) will be equivalent.  
The following result does not use Theorems 3.1—3.3.

**Proposition 3.4** *For an SQP method the statement*

$$\begin{aligned} & \text{(i) (a) } \{B_k\} \text{ and } \{B_k^{-1}\} \text{ are bounded and} \\ & \text{(b) } \lim_{k \rightarrow \infty} \frac{|x_k - x_*|}{|\lambda_k - \lambda_*|} = 0 \end{aligned} \tag{3.14}$$

*implies the statement*

$$\begin{aligned} & \text{(ii) (a) } \{x_k\} \text{ converges } q\text{-superlinearly to } x_* \text{ and} \\ & \text{(b) } \{\lambda_k\} \text{ converges } q\text{-superlinearly to } \lambda_*. \end{aligned}$$

*Proof.* From Theorem 2.1 we see that (i)(a) implies that  $|\lambda_{k+1} - \lambda_*|/|x_k - x_*|$  is bounded uniformly in  $k$ . Hence (i)(b) and (3.13) imply the  $q$ -superlinear convergence of  $\{\lambda_k\}$ .

Moreover, the fact that  $|\lambda_{k+1} - \lambda_*| = O(|x_k - x_*|)$  implies that  $\frac{|\lambda_{k+1} - \lambda_*|}{|x_{k+1} - x_*|} = O\left(\frac{|x_k - x_*|}{|x_{k+1} - x_*|}\right)$ . So

$$\lim_{k \rightarrow \infty} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} = 0 .$$

This proves (ii).  $\square$

Observe that the conditions which allow us to establish  $q$ -superlinear convergence of  $\{\lambda_k\}$ , i.e. (3.11) and (3.14) preclude  $\lambda_k = \lambda_*$  an infinite number of times. Indeed, they do not allow  $\{\lambda_k\}$  to converge too fast relative to  $\{x_k\}$ . In Section 5 we will argue that while this restriction may hold most of the time it is not mathematically realistic.

From Theorem 3.1 we see that the assumption

$$0 < m \leq \frac{|\Delta x_k|}{|\Delta \lambda_k|} \quad \text{for all } k \tag{3.15}$$

is sufficiently strong to make the  $q$ -superlinear convergence of  $\{(x_k, \lambda_k)\}$  equivalent to the condition

$$\lim_{k \rightarrow \infty} \frac{|B_k - A_*| \Delta x_k|}{|\Delta x_k|} = 0. \quad (3.16)$$

Hence, it must be considered somewhat restrictive. We also believe that the condition

$$\lim_{k \rightarrow \infty} \frac{|x_k - x_*|}{|\lambda_k - \lambda_*|} = 0 \quad (3.17)$$

used in Proposition 3.4 is restrictive. It is interesting that (3.15) and (3.17) are incompatible. Specifically, if (3.17) holds and  $\{B_k\}$  and  $\{B_k^{-1}\}$  are bounded, then from Proposition 3.4 we have that  $\{x_k\}$  and  $\{\lambda_k\}$  converge  $q$ -superlinearly. Hence

$$\lim_{k \rightarrow \infty} \frac{|x_{k+1} - x_k|}{|x_k - x_*|} = 1 \quad (3.18)$$

and similarly for  $\{\lambda_k\}$  (see Lemma 8.2.3 of Dennis and Schnabel (1983)). It follows that (3.17) implies

$$\lim_{k \rightarrow \infty} \frac{|\Delta x_k|}{|\Delta \lambda_k|} = 0 \quad (3.19)$$

and (3.15) cannot hold.

## 4 Implications of $(x, \lambda)$ on $x$

Han (1976), (1977), Tapia (1977) and Glad (1978) independently established local and  $q$ -superlinear convergence for the pair  $(x, \lambda)$  for various SQP secant methods as mentioned in Section 1. In general,  $q$ -superlinear convergence for the pair  $(x, \lambda)$  only implies  $r$ -superlinear convergence for  $x$  (or for  $\lambda$ ). However, we will now show that for the SQP quasi-Newton method the  $q$ -superlinear convergence of the pair  $(x, \lambda)$  always implies at least two-step  $q$ -superlinear convergence for  $x$  provided that  $\{B_k\}$  and  $\{B_k^{-1}\}$  are bounded. The result will follow directly from the following lemma.

**Lemma 4.1** *Let  $a_k > 0$ ,  $b_k \geq 0$ ,  $a_{k+1} = O(a_k)$  and  $b_{k+1} = O(a_k)$ . If  $\{(a_k, b_k)\}$  converges to 0  $q$ -superlinearly, then the convergence of  $\{a_k\}$  to 0 is*

at least two-step  $q$ -superlinear. Moreover

$$\lim_{k \rightarrow \infty} \frac{b_{k+1}}{a_{k-1}} = 0.$$

*Proof.* Let  $b_{k+1} = M_k a_k$ . By assumption

$$\lim_{k \rightarrow \infty} \sqrt{\frac{a_{k+1}^2 + b_{k+1}^2}{a_k^2 + b_k^2}} = 0.$$

Hence

$$\lim_{k \rightarrow \infty} \sqrt{\frac{\frac{a_{k+1}^2}{a_{k-1}^2} + \frac{b_{k+1}^2}{a_{k-1}^2}}{\frac{a_k^2}{a_{k-1}^2} + M_{k-1}^2}} = 0. \quad (4.1)$$

Since  $a_k = O(a_{k-1})$  and  $\{M_k\}$  is bounded, it follows from (4.1) that  $\lim_{k \rightarrow \infty} \frac{a_{k+1}}{a_{k-1}} = 0$  and  $\lim_{k \rightarrow \infty} \frac{b_{k+1}}{a_{k-1}} = 0$ .  $\square$

**Theorem 4.1** *For an SQP method the statement*

- (i) (a)  $\{B_k\}$  and  $\{B_k^{-1}\}$  are bounded and
- (b)  $\{(x_k, \lambda_k)\}$  converges  $q$ -superlinearly to  $(x_*, \lambda_*)$

*implies the statement*

- (ii) (a)  $\{x_k\}$  converges to  $x_*$  at least two-step  $q$ -superlinearly and
- (b)  $\lim_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|x_{k-1} - x_*|} = 0$ .

*Proof.* Let  $a_k = |x_k - x_*|$  and  $b_k = |\lambda_k - \lambda_*|$ . From Theorem 2.1 and Corollary 2.1 we have

$$b_{k+1} = O(a_k) \quad \text{and} \quad a_{k+1} = O(a_k).$$

Therefore from Lemma 4.1 we have (ii).  $\square$

## 5 Implications of $(x, \lambda)$ and $x$ on $\lambda$

In this section we establish what we consider to be our main result. We will show that if we have  $q$ -superlinear convergence in the pair  $(x, \lambda)$  and in the variable  $x$ , then we have  $q$ -superlinear or  $q$ -sublinear with unbounded  $q_1$  factor convergence in the multiplier  $\lambda$ . The following lemma is a technical result which will be used later.

**Lemma 5.1** *Let  $\{x_{n_k}\}$  and  $\{\lambda_{n_k}\}$  be subsequences of  $\{x_k\}$  and  $\{\lambda_k\}$ . Assume  $\{B_k^{-1}\}$  is bounded. If*

$$\lim_{k \rightarrow \infty} \frac{|\lambda_{n_k} - \lambda_*|}{|x_{n_k} - x_*|} = 0 ,$$

*then*

$$\lim_{k \rightarrow \infty} \frac{|(I - B_{n_k-1}^{-1} \Gamma_{n_k-1})(x_{n_k-1} - x_*)|}{|x_{n_k} - x_*|} = 1$$

*where  $\Gamma_k$  is defined in Lemma 2.1.*

*Proof.* By Lemma 2.1 we have

$$\frac{x_{n_k} - x_* - (I - B_{n_k-1}^{-1} \Gamma_{n_k-1})(x_{n_k-1} - x_*)}{|x_{n_k} - x_*|} = -B_{n_k-1}^{-1} \nabla h_{n_k-1} \frac{(\lambda_{n_k} - \lambda_*)}{|x_{n_k} - x_*|} . \quad (5.1)$$

By assumption  $\lim_{k \rightarrow \infty} \frac{|\lambda_{n_k} - \lambda_*|}{|x_{n_k} - x_*|} = 0$  and  $\{B_k^{-1}\}$  is bounded. Therefore we have

$$\lim_{k \rightarrow \infty} \frac{|B_{n_k-1}^{-1} \nabla h_{n_k-1}(\lambda_{n_k} - \lambda_*)|}{|x_{n_k} - x_*|} = 0 .$$

It follows from (5.1) that

$$\lim_{k \rightarrow \infty} \frac{|x_{n_k} - x_* - (I - B_{n_k-1}^{-1} \Gamma_{n_k-1})(x_{n_k-1} - x_*)|}{|x_{n_k} - x_*|} = 0 .$$

Therefore

$$\lim_{k \rightarrow \infty} \frac{|(I - B_{n_k-1}^{-1} \Gamma_{n_k-1})(x_{n_k-1} - x_*)|}{|x_{n_k} - x_*|} = 1 . \quad \square$$

**Lemma 5.2** *Let  $Q_k v_k = w_k$  where  $\{Q_k\}$  is a sequence of  $n \times m$  matrices with full column rank and  $\{v_k\}$  is a sequence of vectors. Assume  $v_k \neq 0$  for  $k$  sufficiently large and  $\{Q_k\}$  is bounded and the limit points of  $\{Q_k\}$  have full column rank. If  $\{w_k\}$  converges to 0  $q$ -superlinearly then  $\{v_k\}$  converges to 0  $q$ -superlinearly.*

*Proof.* By the assumption that  $\{Q_k\}$  is bounded and each  $\{Q_k\}$  is of full column rank and the limit points of  $\{Q_k\}$  have full column rank we have

$$m|v_k| \leq |w_k| \leq M|v_k|$$

for some positive constants  $m$  and  $M$  independent of  $k$ . Therefore the fact that  $\{w_k\}$  converges to 0  $q$ -superlinearly will imply  $\{v_k\}$  converges to 0  $q$ -superlinearly.  $\square$

**Theorem 5.1** *If  $\{(x_k, \lambda_k)\}$  and  $\{x_k\}$  converge to  $(x_*, \lambda_*)$  and  $x_*$   $q$ -superlinearly and  $\{B_k\}$  and  $\{B_k^{-1}\}$  are bounded, then either*

- (i)  $\lim_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} = 0$  (i.e.  $\{\lambda_k\}$  converges  $q$ -superlinearly)
- or
- (ii)  $\overline{\lim}_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} = \infty$  (i.e.  $\{\lambda_k\}$  converges  $q$ -sublinearly with unbounded  $q_1$  factor).

Moreover, in both cases  $\lim_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|x_{k+1} - x_*|} = 0$ .

*Proof.* If  $|\lambda_{k+1} - \lambda_*| = O(|\lambda_k - \lambda_*|)$  is not true, then

$$\overline{\lim}_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} = \infty ,$$

and case (ii) of the theorem holds.

Now suppose that  $|\lambda_{k+1} - \lambda_*| = O(|\lambda_k - \lambda_*|)$ . By hypothesis we have

$$\begin{aligned} |\lambda_{k+1} - \lambda_*| &\leq |(x_{k+1} - x_*, \lambda_{k+1} - \lambda_*)| \\ &\leq c_k |(x_k - x_*, \lambda_k - \lambda_*)| \\ &\leq c_k (|x_k - x_*| + |\lambda_k - \lambda_*|) , \end{aligned}$$



where  $\{c_k\}$  converges to 0. It follows that

$$\frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} \leq c_k \left( 1 + \frac{|x_k - x_*|}{|\lambda_k - \lambda_*|} \right). \quad (5.2)$$

Suppose

$$\overline{\lim}_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} = \delta > 0. \quad (5.3)$$

Then there exists a sequence of positive integers  $\{n_k\}$  such that

$$\lim_{k \rightarrow \infty} \frac{|\lambda_{n_k+1} - \lambda_*|}{|\lambda_{n_k} - \lambda_*|} = \delta.$$

Case (i) If

$$\overline{\lim}_{k \rightarrow \infty} \frac{|x_{n_k} - x_*|}{|\lambda_{n_k} - \lambda_*|} < \infty,$$

then from (5.2) we have

$$\lim_{k \rightarrow \infty} \frac{|\lambda_{n_k+1} - \lambda_*|}{|\lambda_{n_k} - \lambda_*|} = 0.$$

However, this contradicts (5.3).

Case(ii) If

$$\overline{\lim}_{k \rightarrow \infty} \frac{|x_{n_k} - x_*|}{|\lambda_{n_k} - \lambda_*|} = \infty,$$

then there exists a subsequence of  $\{n_k\}$ , say  $\{m_k\}$ , such that

$$\lim_{k \rightarrow \infty} \frac{|\lambda_{m_k} - \lambda_*|}{|x_{m_k} - x_*|} = 0. \quad (5.4)$$

From (5.4) and the assumption that  $|\lambda_{k+1} - \lambda_*| = O(|\lambda_k - \lambda_*|)$  we have

$$\lim_{k \rightarrow \infty} \frac{|\lambda_{m_k+1} - \lambda_{m_k}|}{|x_{m_k} - x_*|} = 0. \quad (5.5)$$

Now, since  $\{x_k\}$  converges  $q$ -superlinearly we have from (5.5) and (3.18)

$$\lim_{k \rightarrow \infty} \frac{|\Delta x_{m_k}|}{|\Delta \lambda_{m_k}|} = \infty. \quad (5.6)$$

Theorem 3.1, (5.6) and the fact that we are assuming that  $\{(x_k, \lambda_k)\}$  converges  $q$ -superlinearly imply

$$\lim_{k \rightarrow \infty} \frac{|(B_{m_k} - A_*)\Delta x_{m_k}|}{|\Delta x_{m_k}|} = 0. \quad (5.7)$$

Since  $\{\Gamma_k\}$  defined in Lemma 5.1 converges to  $A_*$  and  $\{x_k\}$  converges  $q$ -superlinearly (see (3.13))

$$\lim_{k \rightarrow \infty} \frac{|(B_{m_k} - \Gamma_{m_k})(x_{m_k} - x_*)|}{|x_{m_k} - x_*|} = 0. \quad (5.8)$$

Let  $w_{k+1} = x_{k+1} - x_* - B_k^{-1}(B_k - \Gamma_k)(x_k - x_*)$  and  $Q_k = -B_k^{-1}\nabla h_k$ . Then from Lemma 2.1 we have

$$w_{k+1} = Q_k(\lambda_{k+1} - \lambda_*). \quad (5.9)$$

Now,

$$\begin{aligned} \frac{|w_{k+1}|}{|w_k|} &= \frac{|x_{k+1} - x_* - B_k^{-1}(B_k - \Gamma_k)(x_k - x_*)|}{|x_k - x_* - B_{k-1}^{-1}(B_{k-1} - \Gamma_{k-1})(x_{k-1} - x_*)|} \\ &= \frac{|\Delta x_k + B_k^{-1}\Gamma_k(x_k - x_*)|}{|\Delta x_{k-1} + B_{k-1}^{-1}\Gamma_{k-1}(x_{k-1} - x_*)|}. \end{aligned}$$

Since  $\{x_k\}$  converges to  $x_*$   $q$ -superlinearly we have

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{|w_{m_k+1}|}{|w_{m_k}|} &= \lim_{k \rightarrow \infty} \frac{|(B_{m_k}^{-1}\Gamma_{m_k} - I)(x_{m_k} - x_*)|}{|(B_{m_k-1}^{-1}\Gamma_{m_k-1} - I)(x_{m_k-1} - x_*)|} \\ &= \lim_{k \rightarrow \infty} \frac{\frac{|(B_{m_k}^{-1}\Gamma_{m_k} - I)(x_{m_k} - x_*)|}{|x_{m_k} - x_*|}}{\frac{|(B_{m_k-1}^{-1}\Gamma_{m_k-1} - I)(x_{m_k-1} - x_*)|}{|x_{m_k} - x_*|}}. \end{aligned} \quad (5.10)$$

From (5.4) and Lemma 5.1 we know that the denominator in (5.10) converges to 1. It follows from (5.8) and the fact that  $\{B_k^{-1}\}$  is bounded and (5.10) that

$$\lim_{k \rightarrow \infty} \frac{|w_{m_k+1}|}{|w_{m_k}|} = 0.$$

Since  $\{B_k\}$  is bounded and  $\nabla h_*$  has full rank the limit points of  $Q_k$  defined by (5.9) have full rank. It follows from Lemma 5.2 and (5.9) that

$$\lim_{k \rightarrow \infty} \frac{|\lambda_{m_k+1} - \lambda_*|}{|\lambda_{m_k} - \lambda_*|} = 0 .$$

This contradicts (5.3).

Both case (i) and case (ii) lead to a contradiction. Hence we must have  $\delta = 0$  and it follows that

$$\lim_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} = 0 ,$$

i.e. case (i) of the theorem holds. The last statement of the theorem follows from Theorem 4.1.  $\square$

Let us end this section by collecting all our results and stating them for the popular SQP secant methods.

**Theorem 5.2** *Consider the SQP Broyden, PSB, DFP or BFGS secant method. In the case of DFP and BFGS assume that the matrix  $A_*$  is positive definite. Then there exist positive numbers  $\epsilon$  and  $\delta$  such that whenever  $|x_0 - x_*| \leq \epsilon$  and  $|B_0 - A_*| \leq \delta$  the iteration sequence  $(x_k, \lambda_k)$  is well-defined and converges to  $(x_*, \lambda_*)$ . In addition we have*

(i)  $\{(x_k, \lambda_k)\}$  converges to  $(x_*, \lambda_*)$   $q$ -superlinearly,

(ii)  $\{x_k\}$  converges to  $x_*$   $q$ -superlinearly,

(iii)  $\lim_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|x_k - x_*|} = 0$ ,

(iv)  $\{\lambda_k\}$  converges to  $\lambda_*$   $q$ -superlinearly or  $q$ -sublinearly with unbounded  $q_1$  factor.

*Proof.* It is known that these secant methods satisfy bounded deterioration and condition (3.5). The bounded deterioration implies that  $\{B_k\}$  and  $\{B_k^{-1}\}$  are bounded. For details see Theorem 3.1, Proposition 4.2 and Corollary 5.5 of Fontecilla, Steihaug and Tapia (1987). The theorem now follows from Theorem 3.1, Theorem 3.3, and Theorem 5.1.  $\square$

## 6 Summary and Concluding Remarks

In Section 3 we exhibited three important characterization theorems and demonstrated that collectively they offer a very powerful analytic tool for studying the convergence properties of SQP methods. We derived several conditions which imply  $q$ -superlinear convergence in  $(x, \lambda)$ ,  $x$  and  $\lambda$  for an SQP method. We also noted that in essentially all applications, the Boggs-Tolle-Wang condition for  $q$ -superlinear convergence in  $x$  has been established by first showing that condition (3.5) holds. Moreover, some of the theory presented in this paper implies that condition (3.5) should be expected to hold. Consequently, we have argued that the Fontecilla-Steihaug-Tapia theorem (Theorem 3.3) which offers a characterization of condition (3.5) is probably a more useful tool than the Boggs-Tolle-Wang characterization theorem (Theorem 3.2).

In general if we have  $q$ -superlinear convergence in the pair  $(x, \lambda)$  we only have  $r$ -superlinear convergence in the variable  $x$  and the multiplier  $\lambda$ . However, Theorem 4.1 shows that if  $\{B_k\}$  and  $\{B_k^{-1}\}$  are bounded, then for an SQP method we always have at least two-step  $q$ -superlinear convergence in the variable  $x$  whenever we have  $q$ -superlinear convergence in the pair  $(x, \lambda)$ .

In Section 5 we showed that the convergence for the multiplier  $\lambda$  was either  $q$ -superlinear or  $q$ -sublinear with unbounded  $q_1$  factor whenever the convergence for the pair  $(x, \lambda)$  and the variable  $x$  were  $q$ -superlinear. We consider this theorem to be the main contribution of the paper. Initially we found this result to be somewhat of a surprise. Indeed, authors have assumed that this convergence is  $q$ -superlinear or at least  $q$ -linear. However, after studying the mechanics of the SQP method we have convinced ourselves that this result should have been expected. Let us now present some discussion along this line.

A highly desirable feature of an iterative procedure is the property that should an iterate happen to coincide with a solution, then the subsequent iterate is also equal to the solution. Clearly, an iterative procedure which lacks this fundamental property cannot have good theoretical  $q$ -convergence behavior. The error could be zero at one iteration and nonzero in the subsequent iteration. This implies that in any analysis which considers the worst case, the  $q_1$ -factor would be unbounded. Even if the error were not zero at any iteration it could be arbitrarily small and one would expect similar statements to hold.

Let us now look at the SQP iterative procedure in terms of  $(x, \lambda)$ ,  $x$  and  $\lambda$  from this point of view. It follows that if  $x_k = x_*$ , then  $\nabla f_k = -\nabla h_k \lambda_*$  and  $h_k = 0$ ; so from (2.1)  $\lambda_{k+1} = \lambda_*$  and from (2.2)  $x_{k+1} = x_*$ . Therefore the establishment of good  $q$ -convergence behavior in  $(x, \lambda)$  and in  $x$  for the SQP method should not be viewed as a complete surprise.

From (2.1) we see that  $\lambda_{k+1}$  does not depend explicitly on  $\lambda_k$ . We should not expect to have  $\lambda_{k+1} = \lambda_*$  whenever  $\lambda_k = \lambda_*$ . Moreover, in most cases there will exist a manifold  $\Omega \in \mathbb{R}^n$  of dimension  $n - m$  such that  $\lambda_{k+1} = \lambda_*$  whenever  $x_k \in \Omega$ . It follows that in the worst-case analysis given by Theorem 5.1 the unbounded  $q_1$ -factor situation is to be expected and cannot be removed from the theorem. The surprise is that Theorem 5.1 says that if the  $q$ -convergence behavior in  $\lambda$  is not arbitrarily bad (unbounded  $q_1$ -factor), then it is essentially optimal ( $q_1$ -factor of zero). It is interesting that both notions are norm independent. We believe that while our numerical experience dictates that in most cases we should expect  $q$ -superlinear convergence in  $\lambda$ , Theorem 5.1 is actually sharp. This means that while the  $q$ -superlinear convergence of  $\{\lambda_k\}$  is likely in any given computation, assumptions which imply it are mathematically restrictive.

It is interesting to point out that the  $r$ -convergence in  $\lambda$  is always superlinear and the unbounded  $q_1$ -factor occurs because the estimate of the multiplier is exceptionally good an infinite number of times.

It is also interesting to point out that in the modified SQP method studied by Tapia and Whitley (1988) if it happens that  $\lambda_k = \lambda_*$ , then the algorithm will converge in the subsequent iteration, i.e.,  $(x_{k+1}, \lambda_{k+1}) = (x_*, \lambda_*)$ . This is due to the very special structure of the eigenvalue problem. Hence it is not unreasonable that they were able to establish the same surprising  $q$ -convergence rate of  $1 + \sqrt{2}$  for the pair  $(x, \lambda)$ , the variable  $x$  and the multiplier  $\lambda$ .

Theorem 5.2 is an up-to-date account of the convergence properties of the SQP Broyden, PSB, DFP and BFGS secant methods. Since these methods satisfy condition (3.5), from Proposition 3.3 we should expect them to give  $q$ -superlinear convergence in  $(x, \lambda)$ ,  $x$  and  $\lambda$ . Theorem 5.2 says that in the unusual case that we do not have  $q$ -superlinear convergence in  $\lambda$ , i.e.  $\lim_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|\lambda_k - \lambda_*|} = \infty$ , we will still have  $\lim_{k \rightarrow \infty} \frac{|\lambda_{k+1} - \lambda_*|}{|x_k - x_*|} = 0$ . Hence, even though it is possible that the  $\lambda$ -sequence may exhibit bad  $q$ -behavior its convergence will be extremely fast. We must conclude that  $q$ -convergence is an

inappropriate and pessimistic measure of convergence for the  $\lambda$ -sequence.

*Acknowledgments.* The authors gratefully thank Richard Byrd and John E. Dennis, Jr. for many helpful suggestions and comments.

## REFERENCES

- BERTSEKAS, D.P. [1982]. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York.
- BIGGS, M.C. [1978]. On the convergence of some constrained minimization algorithms based on recursive quadratic programming. *J. Inst. Math. Appl.*, 21, pp. 67-81.
- BOGGS, P.T., TOLLE, J.W. and WANG, P. [1982]. On the local convergence of quasi-Newton methods for constrained optimization. *SIAM J. Control and Optimization*, 20, pp. 161-171.
- BOGGS, P.T. and TOLLE, J.W. [1985]. An efficient strategy for utilizing a merit function in nonlinear programming algorithms. TR85-5, Curriculum in Operations Research and Systems Analysis, University of North Carolina, Chapel Hill.
- BROYDEN, C.G., DENNIS, J.E., JR. and MORÉ, J.J. [1973]. On the local and superlinear convergence of quasi-Newton methods. *J. Inst. Math. Appl.*, 12, pp. 223-245.
- BYRD, R.H. [1978]. Local convergence of the diagonalized method of multipliers, *J.O.T.A.*, 26, pp. 483-497.
- BYRD, R.H. [1984]. On the convergence of constrained optimization methods with accurate Hessian information on a subspace. TR, CU-CS-270-84, Department of Computer Science, University of Colorado, Boulder, Co.
- BYRD, R.H. [1985]. An example of irregular convergence in some constrained optimization methods that use the projected Hessian. *Math. Prog.*, 32, pp. 232-237.
- BYRD, R.H. and NOCEDAL, J. [1988]. An analysis of reduced Hessian methods for constrained optimization, MCS-P11-1088, Mathematics and Computer Science Division, Argonne National Laboratory.
- COLEMAN, T.F. and CONN, A.R. [1982]. Nonlinear programming via an exact penalty function: Asymptotic analysis. *Math. Prog.*, 24, pp. 123-136.
- COLEMAN, T.F. and CONN, A.R. [1984]. On the local convergence of a quasi-Newton method for the nonlinear programming problem. *SIAM J. Numer.*

- Anal.*, 21, pp. 755-769.
- DENNIS, J.E., JR. [1968]. On Newton-like methods. *Numerisch Mathematik*, 11, pp. 324-330.
- DENNIS, J.E., JR. and MOREÉ, J.J. [1974]. A characterization of  $q$ -superlinear convergence and its applications to quasi-Newton methods. *Math. Comp.*, 28, pp. 549-560.
- DENNIS, J.E., JR. and MOREÉ, J.J. [1977]. Quasi-Newton methods, motivation and theory. *SIAM Review*, 19, pp. 46-89.
- DENNIS, J.E., JR. and WALKER, H.F. [1981]. Convergence theorems for least-change secant update methods. *SIAM J. Numer. Anal.*, 18, pp. 949-987.
- DENNIS, J.E., JR. and SCHNABEL, R.B. [1983]. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, NJ.
- FENYES, P. [1987]. Partitioned quasi-Newton methods for nonlinear equality constrained optimization. Ph.D. dissertation, Department of Computer Science, Cornell University.
- FLETCHER, R. [1981]. *Practical Methods of Optimization, vol. 2, Constrained Optimization*. John Wiley, New York.
- FONTECILLA, R., STEIHAUG, T. and TAPIA, R.A. [1987]. A convergence theory for a class of quasi-Newton methods for constrained optimization. *SIAM J. Numer. Anal.*, 24, pp. 1133-1151.
- FONTECILLA, R. [1988]. Local convergence of secant methods for nonlinear constrained optimization. *SIAM J. Numer. Anal.*, 25, pp. 692-712.
- GABAY, D. [1982]. Reduced quasi-Newton methods with feasibility improvement for nonlinearly constrained optimization. *Math. Prog. Study*, 16, pp. 18-44.
- GARCIA-PALOMARES, U. M. and MANGASARIAN, O.L. [1976]. Superlinearly convergent quasi-Newton algorithms for nonlinearly constrained optimization problems. *Math. Prog. II*, pp. 1-13.
- GILBERT, J.C. [1987]. Maintaining the positive definiteness of the matrices in reduced Hessian methods for equality constrained optimization. IIASA Tech. Rep. WP-87-123, Laxenburg, Austria.
- GILL, P.E., MURRAY, W., SAUNDERS, M.A. and WRIGHT, M.H. [1986]. Some theoretical properties of an augmented Lagrangian merit function. TR SOL 86-6, Department of Operations Research, Stanford University, Stanford, CA.



- GILL, P.E., MURRAY, W., SAUNDERS, M.A., STEWART, G.W. and WRIGHT, M.H. [1985]. Properties of a representation of a basis for the null space. *Math. Prog.*, 33, pp. 172-186.
- GILL, P.E., MURRAY, W. and WRIGHT, M.H. [1981]. *Practical Optimization*. Academic Press, New York.
- GLAD, S.T. [1977]. Properties of updating methods for the multipliers in augmented Lagrangians. *J.O.T.A.*, 28, pp. 135-156.
- GURWITZ, C.B. [1986]. Sequential quadratic programming methods based on approximating a projected Hessian matrix, TR 219, Computer Science Department, New York University, New York.
- HAN, S.P. [1976]. Superlinearly convergent variable metric algorithms for general nonlinear programming problems. *Math Prog.*, 11, pp. 263-282.
- HAN, S.P. [1977]. Dual variable metric algorithms for constrained optimization. *SIAM J. Control and Optimization*, 15, pp. 546-565.
- MURRAY, W. and WRIGHT, M.H. [1978]. Projected Lagrangian methods based on the trajectories of penalty and barrier functions, Systems Optimization Laboratory Report 78-23, Stanford University.
- NOCEDAL, J. and OVERTON, M. [1985]. Projected Hessian updating algorithms for nonlinearly constrained optimization. *SIAM J. Numer. Anal.*, 22, pp. 821-850.
- ORTEGA, J.M. and RHEINBOLDT, W.C. [1970]. *Iterative Solutions of Nonlinear Equations in Several Variables*, Academic Press, New York.
- POWELL, M.J.D. [1978]. The convergence of variable metric methods for nonlinearly constrained optimization calculations. *Nonlinear Programming 3*, Mangasarian, Meyer and Robinson, eds., Academic Press, New York, pp. 27-63.
- POWELL, M.J.D. [1978]. A fast algorithm for nonlinearly constrained optimization calculation, *Lecture Notes in Mathematics*, vol. 630, G.A. Watson, ed., Springer-Verlag, New York.
- STOER, J. [1984]. Principles of sequential quadratic programming methods for solving nonlinear programs. *Computational Mathematical Programming*, K. Schittkowski, ed., Springer-Verlag, New York, pp. 165-207.
- STOER, J. and TAPIA, R.A. [1987]. On the characterization of  $Q$ -superlinear convergence of quasi-Newton methods for constrained optimization. *Math. Comp.*, 49, pp. 581-584.
- TAPIA, R.A. [1974]. A stable approach to Newton's method for general mathemat-

- ical programming problems in  $\mathbb{R}^n$ , *J.O.T.A.*, 14, pp. 453-476.
- TAPIA, R.A. [1977]. Diagonalized multiplier methods and quasi-Newton methods for constrained optimization. *J.O.T.A.*, 22, pp. 135-194.
- TAPIA, R.A. [1978]. Quasi-Newton methods for equality-constrained optimization: Equivalence of existing methods and a new implementation. *Nonlinear Programming 3*, Mangasarian, Meyer and Robinson, eds., Academic Press, New York, pp. 125-164.
- TAPIA, R.A. [1988]. On secant updates for use in general constrained optimization. *Math. Comp.*, 51, pp. 181-202.
- TAPIA, R.A. and WHITLEY, DAVID L. [1988]. The projected Newton method has order  $1 + \sqrt{2}$  for the symmetric eigenvalue problem. *SIAM J. Numer. Anal.*, 25, pp. 1376-1382.
- WILSON, R.B. [1963]. A simplicial algorithm for concave programming. Ph.D. thesis, Harvard University.